

Choice model based on mixed discrete/continuous PDFs

J. Swait, Choice model based on mixed discrete/continuous PDFs, Transportation Research Part B, Vol. 43, Issue. 7, pp.766-783, 2009.

2009/12/16

論文ゼミ #24

原 祐輔

1. Introduction

- この論文ではランダム効用モデルの誤差項を**混合分布**と仮定することで新しいモデルを考える.
- これまで様々なモデルがRUMには提案されている.
 - IIDガンベル分布→MNLモデル(McFadden, 1973)
 - 多変量正規分布→MNPモデル(Daganzo, 1979)
 - GEV分布→NLモデル(McFadden, 1978; Williams, 1977, Ben-Akiva and Lerman, 1985)
 - GNLモデル(Wen and Koppelman, 2001)
 - ordered logit (Small, 1987, 1994)
 - choice set generation logit (Swait, 2001a)
 - choice models based on sums of GEV generation functions (Swait, 2003)
 - dynamic choice models (Swait et al., 2004)
 - 乗数型誤差項モデル(Fosgerau and Bierlaire, 2008)(論文ゼミ#1参照)
 - など

1. Introduction

MNL, MNP, GEV, MXLでさえ一つ的意思決定ルールに従う

効用最大化という意思決定ルール

しかし、人々の意思決定は**複雑**で伝統的なモデルでは必ずしも表現できていないという証拠が知られている(Simon, 1955; Payne et al. 1993)

そこで、本論文では効用がいくつかの離散的な状態(conditionと呼ぶ)にあると仮定することで、あるconditionでは**補償型**の意思決定ルールに従い、別のconditionでは**非補償型**の意思決定ルールに従うと表現することで、意思決定の異質性を表現する。それを本論文では**混合分布**を用いて考える。

混合分布の確率的メカニズムは非補償型意思決定ルールの良い近似でありえるかもしれない。たとえばEBAや辞書編纂型、連結型/分離型などを表現できる。

※連結型…足切り

※分離型…一芸秀逸

RUMへの非補償型ルール of 拡張の既往研究

Gopinath (1995)

SPデータとRPデータにおける意思決定の違いを潜在クラスモデルを用いて表現

Gilbride and Allenby (2004)

選択肢をスクリーニングするために連結型, 分離型, 補償型ルールを特定する
選択モデルを提案

Cascetta and Papola (2000)

Vの中に確率的に選択肢利用可能性を導入することによって暗黙的に選択肢集合の定式化をMNLの中に入れた

Cascetta and Papola (2000)

選択肢集合の定式化において確率的にスクリーニングメカニズムを入れた

2. The 2-condition (2-Mix) model

選択肢*i*の効用を次の式で表す $i \in G$ 選択肢集合

$$U_i = \begin{cases} V_i + \varepsilon_i & \text{with probability } q_i \\ -\infty & \text{with probability } p_i \end{cases} \quad (1)$$
$$p_i + q_i = 1$$

(1)式はある確率*p_i*で選択肢の属性に関わりなく、効用が $-\infty$ となることを表す。上はTradeoff Condition, 下はRejection Conditionとする。

この定義を別の見方をすると、 U_i の確率密度関数が確率 $1-q_i$ で $-\infty$ のパルス波, 確率 q_i で ε_i の確率密度関数に従うと捉えられる。この混合分布は次のように書ける。

$$f_i(U_i) = \underbrace{q_i g_i(V_i, \varepsilon_i)}_{\text{誤差項のpdf}} + (1 - q_i) \underbrace{h_i(-\infty)}_{-\infty \text{にパルスを持つpmf}} \quad (2)$$

誤差項のpdf

$-\infty$ にパルスを持つpmf

(probability mass function: 確率質量関数)

この離散分布と連続分布を結合した混合分布選択モデルの記述を行う

2. The 2-condition (2-Mix) model

$$f_i(U_i) = q_i g_i(V_i, \varepsilon_i) + (1 - q_i) h_i(-\infty) \quad (2\text{再掲})$$

重要な仮定として、
条件確率(condition probability) q_i は効用関数の誤差項 ε_i とは**独立**である。

これは異なる意思決定モードを採用するメカニズムが効用関数の非観測項とは独立であるという仮定と等価。

2.1. Case1 binary choice $G=\{1,2\}$

選択肢1が選ばれる状況は以下の3つの状況があり得る

- (a)両方の選択肢が**Rejection Condition**であり, それは $(1-q_1)(1-q_2)$ の確率で起こる. このとき選択肢1の選択確率はランダムに選ばれるというルールにより単純に $1/2$ である
- (b)選択肢1が**Tradeoff Condition**で選択肢2が**Rejection Condition**である状況は $q_1(1-q_2)$ の確率で起こる. このときの選択肢1の選択確率は1である
- (c)両方の選択肢が**Tradeoff Condition**である状況は q_1q_2 の確率で起こる. このときの選択肢1の選択確率は通常**の補償型のロジットの形で書ける.**

$$P_1 = \frac{(1-q_1)(1-q_2)}{2} + q_1(1-q_2) + q_1q_2 \left(\frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)} \right) \quad (3a)$$

Rejection
Rejection

Tradeoff
Rejection

Tradeoff
Tradeoff

$$P_2 = \frac{(1-q_1)(1-q_2)}{2} + (1-q_1)q_2 + q_1q_2 \left(\frac{\exp(V_2)}{\exp(V_1) + \exp(V_2)} \right) \quad (3b)$$

モデル(3a,3b)がコンセプト的には推定プロセスに抗議票を考慮したWerner (1999)のWTP推定モデルに近い.

2.2. Case2 trinomial choice $G=\{1,2,3\}$

選択肢1が選ばれる状況は先ほどと同様の形で導出できる

$$\begin{aligned} P_1 = & \frac{(1-q_1)(1-q_2)(1-q_3)}{3} + q_1(1-q_2)(1-q_3) \\ & + q_1(1-q_2)q_3 \left(\frac{\exp(V_1)}{\exp(V_1) + \exp(V_3)} \right) + q_1q_2(1-q_3) \left(\frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)} \right) \\ & + q_1q_2q_3 \left(\frac{\exp(V_1)}{\exp(V_1) + \exp(V_2) + \exp(V_3)} \right) \end{aligned} \quad (4)$$

選択肢2, 選択肢3の選択確率も同様に導出できる

2.3. Case3 generalization of the 2-mix model

これらを一般化すると

$$P_i = \frac{\prod_{k \in G} (1 - q_k)}{|G|} + \sum_{C \in \Gamma_i} Q(C) H_{i|C}(V_C, \varepsilon_C) \quad (5a)$$

全てRejection

Tradeoffな選択肢のみでつくられる集合C(iは必ず含まれる)

where

$$Q(C) = \left(\prod_{j \in C} q_j \right) \left(\prod_{j \notin C} (1 - q_j) \right) \quad C \subseteq \Gamma_i \quad (5b)$$

Tradeoffなもの

Rejectionなもの

全選択肢集合Gのうち
iが含まれているもの

ここで ε_C がIIDガンベル分布だとMNLになるので以下のようになる

$$P_i = \frac{\prod_{k \in G} (1 - q_k)}{|G|} + \sum_{C \in \Gamma_i} Q(C) \left(\frac{\exp(V_i)}{\sum_{j \in C} \exp(V_j)} \right) \quad (6)$$

Relationship to Latent class models

潜在クラスモデルは各個人の選好ベクトルによって母集団をS個のグループに分けられると仮定しており、一般的に補償型効用関数に埋め込まれている。本質的にはこれらのモデルは各グループが個別の条件付き選択モデルをもっている緩和することもできる。

よって、潜在クラスモデルはS個の連続的な確率密度関数の混合分布として捉えることが可能である。

$$f_i = \sum_{s=1}^S w_s f_{i|s} \quad (7)$$

セグメントsの
メンバーシップ確率

セグメントsにおける
選択肢iの効用の確率密度関数

潜在クラスモデルのアナロジーで、k-Mixモデルもそれぞれの効用の確率密度関数の連結として書いていると解釈できる。

しかし、コンセプトのモチベーションは大きく異なる。

潜在クラスはマーケットセグメンテーションが目的であったのに対し、混合分布モデルは異なる意思決定メカニズムによる認知プロセスの表現が目的である。

Similarity of 2-mix model to choice set formation models

式(5a)とManski (1977)の2段階確率的選択肢集合生成モデルの類似性について

$$P_i = \frac{\prod_{k \in G} (1 - q_k)}{|G|} + \sum_{C \in \Gamma_i} Q(C) H_{i|C}(V_C, \varepsilon_C) \quad (5a) \text{再掲}$$

$$P_i = \sum_{C \in G} P_{i|C} \tilde{Q}(C) \quad (8)$$

選択肢は確率 p_i で独立に利用可能とする
Probabilistic Independent Availability Logit
(PIAL)modelとする (Swait, 1984)

$$\tilde{Q}(C) = \frac{\left(\prod_{j \in C} \rho_j \right) \left(\prod_{j \notin C} (1 - \rho_j) \right)}{1 - \left(\prod_{j \in G} (1 - \rho_j) \right)} \quad C \subseteq G \quad (9a)$$

$$P_{i|C} = \frac{\exp(V_i)}{\sum_{j \in C} \exp(V_j)} \quad (9b)$$

2-mix モデルは非補償型効用関数を入れることに着眼しており
選択肢集合生成モデルは選択肢集合の中に入っている選択肢の選別に着眼している

ただし、2-mixモデルも選択肢集合生成モデルもGが大きくなるにつれて
次元の呪いが発生することは共通している。

3. The 3-condition (3-Mix) model

- (a) **Tradeoff Condition** : 効用は属性に依存し, 補償型である
- (b) **Rejection Condition** : 効用は常に $-\infty$
- (c) **Dominance Condition** : 効用は常に $+\infty$

$$s_i = \begin{cases} +1 & \text{if } i \in G \text{ is in Dominance Condition} \\ 0 & \text{if } i \in G \text{ is in Tradeoff Condition} \\ -1 & \text{if } i \in G \text{ is in Rejection Condition} \end{cases} \quad (11)$$

$$U_i = \begin{cases} +\infty & \text{if } s_i = +1 \\ V_i + \varepsilon_i & \text{if } s_i = 0 \\ -\infty & \text{if } s_i = -1 \end{cases} \quad (12)$$

$$\Pr(s_i) = \begin{cases} p_i & \text{if } s_i = +1 \\ q_i & \text{if } s_i = 0 \\ r_i & \text{if } s_i = -1 \end{cases} \quad (13)$$

$$p_i + q_i + r_i = 1$$

3. The 3-condition (3-Mix) model

Ω : 選択肢集合 $|G|$ のありうる状態

ω_l : Ω の構成要素 $\omega_l = \{s_1, \dots, s_{|G|}\} \quad l = 1, \dots, 3^{|G|}$

ここで $Q(\omega_l)$ は ω_l の状態が起こる確率であり

$$Q(\omega_l) = \underbrace{\left(\prod_{j \in G: s_j = +1} p_j \right)}_{\text{Dominance}} \underbrace{\left(\prod_{j \in G: s_j = 0} q_j \right)}_{\text{Tradeoff}} \underbrace{\left(\prod_{j \in G: s_j = -1} (1 - p_j - q_j) \right)}_{\text{Rejection}}, l = 1, \dots, 3^{|G|} \quad (14)$$

選択肢 i の選択確率は膨大な状態 ω の条件付き確率から計算される。

- (a) 選択肢 i が **Dominance Condition** の状態にあるとき、選択確率は同じ状態にある選択肢の中からランダムに選ばれる
- (b) すべての選択肢が **Rejection Condition** の状態にあるとき、選択確率はランダムに選ばれる
- (c) 選択肢 i が **Tradeoff Condition** にあり、Dominance Condition な選択肢が存在しないとき、補償型の効用によって選択確率が決定する

3. The 3-condition (3-Mix) model

選択肢*i*の選択確率は

$$P_i = \sum_{l=1}^{3^{|G|}} R(i | \omega_l) Q(\omega_l) \quad (15)$$

$$R(i | \omega_l) = \begin{cases} \left(\sum_{j \in G} I(s_{ij} = +1) \right)^{-1} & \text{if } s_{ii} = +1 \\ H_{i|c}, C \subseteq G \text{ s.t. } s_{lk} \neq +1, \forall k \in C & \text{if } s_{ii} = 0, s_{ij} = 0 \text{ at least one } j \neq i, \\ & \text{and } s_{ij} \neq +1, \forall j \in G \\ 1 & \text{if } s_{ii} = 0, s_{ij} = -1, \forall j \neq i, j \in G \\ |G|^{-1} & \text{if } s_{ij} = -1, \forall j \in G \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

ここで $I(\cdot)$ は指示関数でそうであれば1となる関数.
Hは補償型モデル

3. Discussion of 3-Mix model

- 基本的には2-Mixモデルと同様
- 一番の違いはDominance Conditionがあること
- Simon (1955)は人々は効用最大化ではなく、満足化ルールで意思決定をしていると仮定している。つまり、効用最大な選択肢を探す認知コストを考え、あるレベルを越えているものを一回発見すればそれ以上探索しない。
- これはDominance Conditionによって表現可能であるかもしれない

4. Application of proposed models

- データ
 - 通勤手段RPデータ
 - サンパウロ都市圏における2826のhome-based work trip
 - 1987 O/D surveyのデータ
 - 分担率は自動車運転：20.3%，自動車送迎：4.5%，バス：40.6%，地下鉄18.6%，企業バス：5.5%，徒歩・自転車：9.0%，バイク：1.6%

Table 1
Work trip mode choice models for São Paulo.

	MNL model	2-Mix model ^a	3-Mix model ^b	PIAL model ^c
Utility function	(asymptotic t-stats)			
<i>Alternative-specific constants</i>				
Auto drive	3.138 (16.7)	3.462 (13.5)	26.603 (23.2)	3.489 (13.5)
Auto passenger	0.540 (2.5)	0.806 (2.1)	25.380 (25.2)	0.677 (2.9)
Bus	3.307 (18.7)	3.795 (19.7)	38.578 (18.1)	3.568 (15.9)
Metro	3.662 (19.8)	3.988 (16.1)	35.301 (22.7)	3.743 (17.7)
Company bus	1.094 (6.0)	1.068 (5.5)	28.896 (0.0)	1.274 (3.6)
Walk/Bicycle	1.098 (6.5)	7.143 (3.8)	-9.352 (-0.0)	6.927 (2.8)
Motorcycle	-0-	-0-	-0-	-0-
<i>Impedances</i>				
Travel time (min)/100	-0.008 (-4.6)	-0.021 (-9.8)	-0.092 (-3.9)	-0.022 (-9.1)
Cost/income	-0.018 (-0.5)	-0.074 (-3.4)	0.121 (0.6)	-0.278 (-4.1)
<i>Socio-demographics</i>				
Income (auto drive)	0.022 (5.6)	0.031 (4.1)	0.202 (3.4)	0.022 (3.3)
Income (auto pass)	0.030 (5.2)	0.038 (4.8)	0.193 (3.7)	0.027 (4.2)
<i>Tradeoff condition parameters</i>				
Auto drive	-	1.980 (3.6)	11.070 (0.0)	
Auto passenger	-	1.362 (1.0)	16.169 (0.0)	
Bus	-	28.896 (0.0)	2.185 (10.2)	
Metro	-	13.368 (0.0)	12.533 (0.0)	
Company bus	-	23.020 (0.0)	28.643 (0.0)	
Walk/Bicycle	-	-2.299 (-23.7)	-11.695 (-0.0)	
Motorcycle	-	20.012 (0.0)	12.423 (0.0)	
<i>Dominance condition parameters</i>				
Auto drive			11.098 (0.0)	
Auto passenger			13.168 (69.8)	
Bus			-10.805 (0.0)	
Metro			11.947 (0.0)	
Company bus			25.112 (83.2)	
Walk/Bicycle			-2.119 (-24.0)	
Motorcycle			8.656 (46.8)	
<i>Independent availability parameters</i>				
<i>Alternative-specific constants</i>				
Auto drive	-	-	-	
Auto passenger	-	-	-	
Bus	-	-	-	
Metro	-	-	-	
Company bus	-	-	-	
Walk/Bicycle	-	-	-	
Motorcycle	-	-	-	14.187 (0.0)
<i>Goodness-of-fit</i>				
LL(0)	-4298.81	-4298.81	-4298.81	-4298.81
LL(convergence)	-2798.33	-2737.62	-2698.90	-2735.01
Number free param.	10	17	24	17
ρ^2	0.3490	0.3632	0.3722	0.3638
ρ^2 Akaike	0.3467	0.3599	0.368	0.3605

^a In the 2-Mix model, $q_i = [1 + \exp(-\delta_i)]^{-1}$, where δ is the parameter reported in table.

^b In the 3-Mix model, $p_i = \exp(\tau_i)/D$, $q_i = \exp(\gamma_i)/D$, $r_i = 1/D$, where $D = \exp(\tau_i) + \exp(\gamma_i) + 1$, where τ , γ are reported in table.

^c In the PIAL model, $\rho_i = [1 + \exp(-\lambda_i)]^{-1}$, where λ is the reported parameter value.

Tradeoffの確率

$$q_i = \frac{\exp(\gamma_i)}{\exp(\tau_i) + \exp(\gamma_i) + 1}$$

Tradeoffの確率

$$q_i = \frac{1}{1 + \exp(-\delta_i)} = \frac{1}{1 + \exp(-1.98)} = 0.88$$

Dominanceの確率

$$p_i = \frac{\exp(\tau_i)}{\exp(\tau_i) + \exp(\gamma_i) + 1}$$

Rejectionの確率

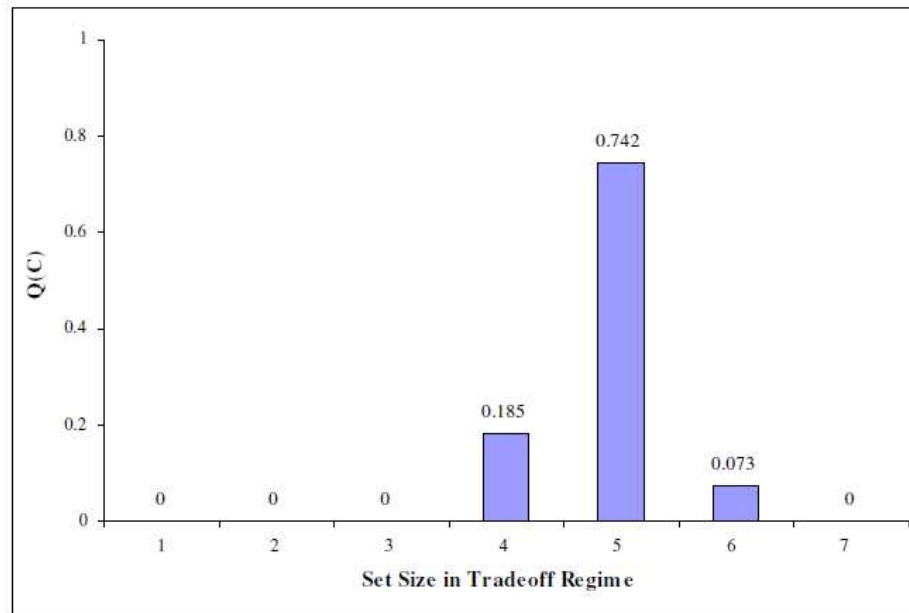
$$r_i = \frac{1}{\exp(\tau_i) + \exp(\gamma_i) + 1}$$

各選択肢のcondition probability

Table 2
São Paulo *k*-Mix models: estimated alternative-specific condition probabilities.

2-Mix/3-Mix			
Alternative	Pr[Dominance]	Pr[Tradeoff]	Pr[Rejection]
Auto drive	0.00/0.51	0.88/0.49	0.12/0.00
Auto passenger	0.00/0.05	0.80/0.95	0.20/0.00
Bus	0.00/0.00	1.00/0.90	0.00/0.10
Metro	0.00/0.36	1.00/0.64	0.00/0.00
Company bus	0.00/0.03	1.00/0.97	0.00/0.00
Walk/Bicycle	0.00/0.11	0.09/0.00	0.91/0.89
Motorcycle	0.00/0.02	1.00/0.98	0.00/0.00

(a) 2-Mix Model



選択肢集合が7つ
 0.064
 選択肢集合が6つ
 0.661
 ?????謎?

4. Application of proposed models 2

- データ 2

- 娯楽時キャンプ場選択のSPデータ
- 選択肢は6つ（国立公園）
- キャンプサイト 1, 2は実験計画法でコントロールされた属性
- キャンプサイト 3は古典的なキャンプサイトの属性で固定しているが値段だけ変化する
- キャンプサイト 4は国立公園の外にある
- キャンプサイト 5は国立公園内の町のホテルやロッジに泊まる
- キャンプサイト 6は家にいる
- キャンプサイト 4～6は細かな属性はなく、これだけ書いてある
- データ数は1776

Table 3
Recreation campsite selection models.

	MNL model	2-Mix model ^a
Utility function		
<i>ASCs</i>		
Campsite 1	2.492 (27.3)	2.565 (1.4)
Campsite 2	2.517 (27.9)	2.549 (1.4)
Campsite 3	2.077 (11.9)	1.962 (1.1)
Camp Out of Park	1.071 (11.9)	0.066 (0.1)
Hotel/Lodge/Cabin	-0.479 (-3.6)	-0.269 (-0.2)
Stay Home	-0-	-0-
<i>Attributes (Campsites 1,2,3)</i>		
<i>Hookups</i>		
Water/Elec/Sewer	-0.687 (-12.1)	3.716 (1.1)
Electricity	0.087 (1.5)	-1.644 (-1.0)
None	-0-	-0-
<i>Fees</i>		
Overnight L	-0.204 (-10.0)	-0.303 (-5.8)
Q	-0.125 (-2.5)	-0.118 (-1.4)
Hookup - WES	-0.432 (-4.5)	-0.492 (-0.3)
Hookup-Elec	-0.299 (-3.7)	-0.175 (-0.6)
<i>Other</i>		
Visual separation	0.377 (10.1)	0.523 (6.8)
Park staff present	0.002 (0.1)	-0.032 (-0.4)
Security patrols	0.153 (3.9)	0.151 (2.4)
Showers	0.465 (13.9)	0.748 (7.5)
Drinking water	-0.031 (-0.8)	-0.008 (-0.1)
Toilet type 1	0.240 (4.4)	0.148 (1.5)
Toilet type 2	0.153 (2.8)	0.376 (3.5)
Toilet none	-0-	-0-
Cultural programs	0.043 (1.1)	0.008 (0.1)
Location > 10 km from town	0.057 (1.4)	0.078 (1.3)
Tradeoff condition function		
<i>ASCs</i>		
Campsite 1	-	1.741 (0.2)
Campsite 2	-	1.828 (0.2)
Campsite 3	-	0.920 (0.1)
Camp out of park	-	-0.971 (-3.1)
Hotel/Lodge/Cabin	-	-2.511 (-3.3)
Stay home	-	-2.164 (-4.3)
<i>Attributes (campsites 1,2,3)</i>		
<i>Hookups</i>		
Water/Elec/Sewer	-	-2.835 (-0.4)
Electricity	-	-1.136 (-0.2)
None	-	-0-
<i>Fees</i>		
Overnight L	-	-0.945 (-0.1)
Q	-	-0.137 (-0.0)
Hookup-WES	-	-0.377 (-3.2)
Hookup-Elec	-	-0.375 (-1.5)
<i>Other</i>		
Visual separation	-	0.206 (2.1)
Park staff present	-	-0.077 (-0.9)
Security patrols	-	0.199 (2.3)
Showers	-	0.318 (3.1)
Drinking water	-	-0.075 (-0.9)
Toilet type 1	-	0.133 (1.0)
Toilet type 2	-	0.347 (3.1)
Toilet none	-	-0-
Cultural programs	-	-0.069 (-0.7)
Location > 10 km from town	-	0.170 (1.8)
LL(0)	-25004.00	-25004.00
LL(convergence)	-16537.75	-16198.80
Number free param.	20	41
ρ^2	0.339	0.352
ρ^2 Akaike	0.338	0.351

^a In the 2-Mix model, $q_i = [1 + \exp(-\delta'Z_i)]^{-1}$, where δ is the parameter vector reported in table and Z_i is the vector of ASC's and attributes.

$$q_i = \frac{1}{1 + \exp(-\delta_i)} = \frac{1}{1 + \exp(-1.98)} = 0.88$$

$$q_i = \frac{1}{1 + \exp(-\delta'X_i)}$$

選択肢1, 2のQ(C)を
各変数で構造化している

	変数	パラメータ	t値	パラメータ	t値
利用	利用権日時にMM訪問	0.860	3.12	18.740	0.01
	男性ダミー	1.722	6.81	6.725	3.76
	log(距離)	-1.392	-3.86	-6.400	-0.44
	休日ダミー	2.918	3.60	21.556	0.09
販売	log(距離)	-0.405	-2.75	-2.986	-3.68
	平日ダミー	3.335	6.34	23.227	0.10
	午前ダミー	-0.476	-2.09	-3.539	-2.74
	スケジュール	0.198	1.63	0.257	0.69
放置	選択肢固有定数	4.572	6.94	40.058	0.03
	スケジュール	0.298	2.49	0.650	1.62
	40代ダミー	0.288	5.54	19.988	0.01
Tradeoff Condition	利用	Tradeoffの確率 利用 0.465	Rejectionの確率 利用 0.535	-0.139	-0.30
	販売	販売 0.995	販売 0.005	5.344	1.13
	無視	無視 0.955	無視 0.045	3.047	22.74
	初期対数尤度	-1535.860		-1535.860	
	最終対数尤度	-311.508		-295.151	
	修正済み適合度	0.790		0.799	